HMMER Sequences in the Pol Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_HXB2R	HIVHXB2R	K03455	Starcich,B.	Science 227, 484 (1985)
D_ELI	HIVELI	KO3454	Alizon,M.	Cell 46, 63 (1986)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
CPZANT	SIVCPZANT	U42720	Vanden Haesevelde,M.	Virology 221, 346 (1996)
A_ROD	HIV2ROD	M15390	Clavel,F.	Nature 324, 691 (1986)
B_EHOA	HIV2EHOA	U27200	Rey-Cuille,M.A.	Virology 202, 471 (1994)
SD_MM251	SIVMM251	M19499	Franchini,G.	Nature 328, 539 (1994)
STM_STM	SIVSTM	M83293	Novembre,F.J.	Virology 186, 783 (1992)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
TAN_AGMT17	SIVAGMT17	L19254	Hirsch,V.M.	Virology 197, 426 (1993)
SAB_SAB1C	SIVSAB1C	U04005	Jin,M.J.	EMBO J. 13, 2935 (1994)
SYK_SYK	SIVSYK	L06042	Hirsch,V.M.	JVI 67, 1517 (1993)

HIV1 POL

The following alignment and most likely sequence were generated using the HMMER program as described in Part III. For simplicity, only representative types and subtypes are shown. Annotation is based on HIV1s, therefore the user should be cautious about its applicability to other PIV sequences. Five key sites for drug resistance are indicated; for further information regarding drug resistance, see the section by J. Mellors and colleagues in Part III. Cleavage sites are indicated by '\/'. Ten conserved, gapless blocks were identified using the BLOCKMAKER and Motif programs; these are shown by shading. The Motif "Cobbler" sequence follows this alignment.

```
most-likely                                          FFRE.DLAFPQGK.AAEQFPR........     19
A_U455                                               ----.N---Q--E.-R-F...........     16
B_HXB2R                                              ----.----L---.-R-F...........     16
D_ELI                                                ----.N-------.-G-L...........     16
O_ANT70                                              ---Q.I--S.G-H.E-R-L..........     16
O_MVP5180                                            ----.V--S.G-H.E-R-L..........     16
CPZGAB                                               ----.R-----RE.-RQL...........     16
CPZANT                                               ----T-PHV.V-V.QTRELCA........     19
A_ROD                                             TGR---T.GPL...--.E-P-L--........     19
B_EHOACG       KAGMLEMWTART..................HHVKMPRKTGG---V.RPL...--.E-S----PGTPGDSAI     48
D_MM251        VLELWEGGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGG---P.WSM...--.E-P---H........     57
STM_STM     RPEKTRLLEMWTTGP..................SDGQMPRKTGG---L.WPM...--.E-P-L-H........     42
VER_AGM3    T.KKDEMLEMWE.TR..................AFSKRLQRTGK---V.WPVD..-S.ETQK-S-........     41
GRI_AGM677     VRQNWPYGK.......................RLQEWTGK---V.WPL...-RSETKK-CAIQ.......     36
TAN_AGMT17                                     RRTGQ--K-.NARTLRFQ.--QF-RARR.......     26
SAB_SAB1C                                            ---V.WPLG..QR.ETQE-...........     15
SYK_SYK                                              ----.RMGG.LQE.-PS-L-SEERAICSPS     27

most-likely ................GPSSSGQARANSPTRRESQ.......................................     38
A_U455      ....................--E-T------S-NLWD.GGKDD.LPCETG................AER..Q.GT     50
B_HXB2R     ....................--E-T---------L-V.WGRDNNSPSEAG................ADR..Q.GT     51
D_ELI       ....................-PK-T------S--LRV.WGRDN.PLSKTG................AER..Q.GT     50
O_ANT70     ....................CAETSTPI---DGGGS..........EGTGES.............GTERGPE.RA     47
O_MVP5180   ....................CAETSVPI---NGGGS...EGTRESESEGG................SGRA.....     47
CPZGAB      ....................CAE-N-T-G--D--LWVPGGR.....EEPGE..............ERGR..E.QS     49
CPZANT      ................EGG---..---S-H-DLS...GGAQ.EDSEGS................QGG..G.GT     51
A_ROD       ................----A-.-DT--TPSGS-S...GSTG.EIYAARE..............KTERA.ERET     57
B_EHOACG    ........CAPDEPSIRHDT--.CDSICTPC-S-R...GDAK.ELHATRE..............EAEGE.QRET     94
D_MM251     ................-S-A--.-D--CSP-GP-C...GSAK.ELHAVGQ..............AAERK.QREA     95
STM_STM     ................--NT--.-D--C-S--P-C...GSAE.KLHAAGQ..............EAERE.QEDT     80
VER_AGM3    ................RY-WG-...--CAPST--I...RPCK.EAPAAI..................CR..QGEA     73
GRI_AGM677  ...............RRH-W--...T---PNGN-L...RSSK.EAPPAVCREGTAPERGERTDKETEG..E.RS     85
TAN_AGMT17  ...............SRSDG-PDR--WI-H-SA-C...RSLRSSQETPGAVCKKR.....RPTKEAKR..E.GT     74
SAB_SAB1C   ....................P-DLHQT--SPNGTGL.........QQAGGK..............LVCR..Q.TS     44
SYK_SYK     GGHRRWAMAHMVSTDEPT-T-E................GTEL.SLEETPHQQG.........SALAEG..E.QW     72

                                                                   <- gag cds end
                                                        \/ protease
most-likely ......................................VGFAFPQITLWQRPLVTAKIEGQLVEALLDTGADD     73
A_U455      ......................................DS-S-------------V--G---I----------     85
B_HXB2R     .......................................-S-N---V---------I--G---K----------     86
D_ELI       .......................................-S-N------------AI--G---K----------     85
O_ANT70     ......................................LSVCL---P--D--I---RVG-H-C-V--------     82
O_MVP5180   .......................................-PICL---P--D--I----VG-H-C----------     82
CPZGAB      ......................................ISTNL----------IPV-V----C----------     84
CPZANT      ......................................TSLV--E-P-----MMEVL-Q--KCQ---------     86
A_ROD       IQGSDRGLTAPR................AGGDTIQGATNR-L-A--FS--K--V---Y----P--V--------    115
B_EHOACG    LQG..................................GDR---A--FS--R--V-K-T----S--V--------    134
D_MM251     LQG..................................GDR---A--FS--R--V---H----P--V--------    135
STM_STM     LQG..................................GDR---A--FS--R--V---H----P--V--------    120
VER_AGM3    VEGTKEKTT......................SSESRLDR-IF-ELP.--R--IK-VY---VPIR---------    123
GRI_AGM677  ......................................G-CFLELP.--R--MKRVI---TP-Q---------    119
TAN_AGMT17  ......................................G-LFLEF                                 81
SAB_SAB1C   DQRTRARRSSNSPVKAVCCSGETAETAVAKPLATTEPLRG-LQL--VS--R--MK-VY----K-T---------    118
SYK_SYK     EGG..................................DQESLSLS-LS--R--MIEVDV--D--QM-V------    112
```

```
most-likely TV..EEGIL..L.GR.WKPKIIGGIGGFIKVKEYDQVEIEIC...G..KRAIGTVLVGPTPVNIIGRNLLTQLG    136
A_U455      --L.-DIN-..P.-K.---------------RQ---IL----...-..-KT-----------------M---I-    149
B_HXB2R     --L.--MS-..P.--.----M----------RQ---IL----...-H.-.----------------------I-    150
D_ELI       --L.--MN-..P.-K.----M----------RQ---IP----...-Q.-.----------------------I-    149
O_ANT70     --..LNN-Q..-E-K.----M--------------N-TV--E...-..REVQ----------------I--G--    146
O_MVP5180   --..LNN-Q..-E--.-T--M-------------NN-TV-VQ...-..-EVQ------------L---I--G--    146
CPZGAB      --..I-R-Q..-Q-L.----M-----------QF-N-H---E...-R.-.VV----------------I-----    148
CPZANT      --..V---H..-Q-N.----T--------S-QQ-NK-P-Q-G...D..RTVLA---L--N--------V-CL--    150
A_ROD       SI..VA--E..-.-NNYS---V-------NT---KN----VL...N..-KVRA-IMT-D--I--F---I--A--    179
B_EHOACG    SI..VA--E..-.-SNYT---V-------NTN--KN----VV...-..--VRA--MT-D--I--F---I-NS--    198
D_MM251     SI..VT--E..-.-PHYT---V-------NT---KN-K--VL...-..--IK--IMT-D--I--F------A--    199
STM_STM     SI..VA--E..-.-LQYT--VV-------NT--FKN-N--VL...-..-KIK--IMT-D--I--F------A--    184
VER_AGM3    -IIK-ADLQ..-S-T.-----------GLN----SDR-VRLE...D..-ILR--I-I-S--I------I-APA-    189
GRI_AGM677  -IIQ-KDLHFPP.HKP-RS-VV-----G-H----QG-QVQLE...D..-IIT-SI-I-S--I------I-A-A-    187
SAB_SAB1C   S-..IQ--E..-.-DN---R-------C-N--A-HNQ-VK-E...D..-TCKA-I---E---------V-A---    182
SYK_SYK     -IIR--D-Q..-.HQP-S---V--L--N-T-RQ-RNIRFTVVKPS-KR-QVE--L---------L---I--K--    183

       protease \/ p66,p51                                   |M41L
most-likely CTLNFPISP..IETVP.VKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAIKKK    207
A_U455      ---------..-----.-----E-------------------T---N---------------------------    220
B_HXB2R     ---------..-----.------------------------------------------------------------    221
D_ELI       ---------..-----.-------------------------T----D--------R----------I------    220
O_ANT70     ---------..-AP--.-----------------SK---E--TA--Q---Q-----R----------I------    217
O_MVP5180   ---------..-AP--.-----------------SR---E--TA--Q---Q-----R----------I------    217
CPZGAB      ---V----S..-----.-----------------SA------T---Q--------------------I------    219
CPZANT      --------K..V----.----E-----R------SK---E--K---DKL-A-N---R---D------I------    221
A_ROD       MS--L-VAK..V-PIK.IM----K----LR-----K---E--K---EK-----QLEEAP-T------T------    250
B_EHOACG    M-----VAR..--P-K.-Q---EK----IR----SK---L--K---EK-----QLEEAP-T----S-T------    269
D_MM251     MS--L--AK..V-P-K.-T----KV---L-----SK---V--R---EK---D-QLEEAP-T------T------    270
STM_STM     MS----VAK..V-P-K.-T----K----I-----SK---E--K---EK---D-QLEEAP-T------T------    255
VER_AGM3    AK-VMGQLSE.QIPITP----E-AR--FL-----SK------Q---DQL----------G--A------C----    262
GRI_AGM677  MK-VMGVLSSQ--ETK.-Q--E-K----L-----SR---E--T---KQ--E---L-R--G--------------    260
SAB_SAB1C   V---LTQRE..--PIK.-H----Q---RIR----SK---E--KA--EDL--Q-HLER--------------R--    253
SYK_SYK     VK-VMVQTA..L-P-K.-S---DKEL-RL-----SV--LE--KA-VED-L-A-QLE-AS-T--------V-R--    254

               |K70R
            |D67N
most-likely DSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIPSVNNETP    281
A_U455      ------------------------------T-------------------------S-----------I-----    294
B_HXB2R     ------------------------------------------------------------------I-----    295
D_ELI       -------------------------------------------------------------S-I-----    294
O_ANT70     -G-----------------E-----------G---Q-Q-------------C---P------------------    291
O_MVP5180   -------------------------------G---QRQ-------------C---P------------------    291
CPZGAB      ---------------------------------------------------C---K------------I-----    293
CPZANT      -TS--------------------I-----------Q---------------I---Q------------------    295
A_ROD       -KN---M-I-------V----T-I----------A--RRI-----------I--H----P-----L-----AE-    324
B_EHOACG    -KN---M-I-------V--E-T------------AS--RI---------------P---Q-----L-A---AE-    343
D_MM251     -KN---M-I------RV----T------------A-R-RI----I------I----E--Q-----L-----AE-    344
STM_STM     -KN---M-I-------V----T-I----------A-RRRI-----------I----G--Q-----L-----AE-    329
VER_AGM3    -KSQ--M---------A----F----------S-FE-MTEI----I----Y-I---PE--------------QG-    336
GRI_AGM677  -K-Q--M---------A----F------------Q---QI--I-I----Y-I--CKE--------------TG-    334
SAB_SAB1C   -K-Q--I-M---Q---S----Q------------QQREQI----I------C---P--Q------------RE-    327
SYK_SYK     -KK---M-I---K--EA----F----------G--.-RQKL-II-LK---Y-----KE--P-----V--I--AS-    327
```

HIV1 POL

```
                                                   polymerase
                                                    <-motif->
most-likely GIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIVIYQYMDDLYVGSDRLEIGQHRTKIEELREHLLRWG    355
A_U455      -V---------------S--------------S-H-----------------.-------A------A---S--    367
B_HXB2R     ------------------------------------------------------.--------------Q------    368
D_ELI       -------------------------------------EM--------------.-----------K---------    367
O_ANT70     ----------------------------D---RD--ELE-C-----------.-PLTE--KR--L-----YQ--    364
O_MVP5180   -V--------------------------D----S--EVE----I--------.-PLAE--KRV-L-----YQ--    364
CPZGAB      -V---------------S--------------EK----T-------------.---D---K-V----Q---K--    366
CPZANT      ----C----------------A------A---DKY-AVE-------------.M--TA--EM--K--Q--QV--    368
A_ROD       -K--I-K--------------HT-RQV------A-K-VI-I-----ILIA---TDLEHD-V.VLQ-K-L-NGL-    397
B_EHOACG    -K--L-K--------------YT-A-V-D----A-N-VT-I-----IL-A---SDLEHD-V.VSQ-K-L-NNM-    416
D_MM251     -K--I-K--------------YT-RHV------A---VTLV-----ILIA---TDLEHD-V.VLQ-K-L-NSI-    417
STM_STM     -K--I-K--------------YT-RN------RA---VTLI-----ILIA---TDLEHD-V.VLQ-K-L-NNL-    402
VER_AGM3    -T---F-C---------T---NTAAS---EIK-ELKPLT-V------W---QED-YTHD-L.V-Q--MK-SA--    409
GRI_AGM677  -----F-C---------T---NTAAN---EIKRHT-GLE-V------WLA--HD-TRHNQQ.VDIV-KM--EK-    407
SAB_SAB1C   ------K----------T---TTAN---QE--QK---VD-------MLIA---PKAEHLVM.VQQ--DY-ET--    400
SYK_SYK     -E---FT----------T----TINQ--Q----KYS-LTLI------LI-T--S-KAHQEI.VQQIVTA--KV-    400

                 |K219Q
             |T215Y
most-likely FTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPE.KDS.WTVNDIQKLVGKLNWASQIYPGIKVKQLCKLLRG    427
A_U455      -I-----------------------------Q---.---.--------------------A-------------    439
B_HXB2R     L----------------------------------.---.--------------------------R--------    440
D_ELI       --R--------------------------S-K---.-E-.-------N--ER-------------R--------    439
O_ANT70     ----------------------------S-Q--N.--V.---------I----------Q--R-RE----I--    436
O_MVP5180   -------------------------------Q--D.-EV.--------------------Q--R--E----I--    436
CPZGAB      -------------------------------Q---.-EV.---------I--------------I------I--    438
CPZANT      LE------------Q--------------K-K---.P-D.------------------------T----R-I--    440
A_ROD       -S---E-F--D--YH------W-T--KL-K-Q--Q.-EI.-----------V----A-L-----T-H--R-I--    469
B_EHOACG    -S--EE-F--D---K------W-K--KL-K-Q---.-EV.-----------V----A-LF----TRHI---I--    488
D_MM251     -S--EE-F--D---Q------W-T--KL-K-E--Q.RET.-----------V----A-------T-H--R-I--    489
STM_STM     -S--EE-F--D---Q------W-T--KL-K-E--Q.R-V.-----------V----A-------T-H----I--    474
VER_AGM3    LE-----V--K--YE----K-W-H--QISS-E-ED.-EE.-------R--------A-L---LRT-N----I--    481
GRI_AGM677  LE-----V-R---WE----K---N---INK-E--P.LEGE----K---V--V------------T-HT-AM---    480
SAB_SAB1C   -K--E--F--D--Y-------Y-K--QL-E-T---.REE.--------------------T---T-H--R-I--    472
SYK_SYK     -KV-KE-W-DQY-MQ-L--T------QL-K-E--NID-E.I---QL---I-V--------S---T-E---CI--    473

most-likely TKA.LTEVVPLTEEAELELAENREI...LKQEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEPF.KNLKTGK    496
A_U455      A--.--DI-T----------------...--.D-------------V--------D-----------.-------    507
B_HXB2R     ---.----I----------------...--.------------------------------------.-------    508
D_ELI       ---.----I----------------...--.----------------------H------------.-------    507
O_ANT70     --S.--------SR------E----R...--.Q-------Q-D---WVN-----GE---------EH.-------    504
O_MVP5180   --S.--------SK------E----K...--.--------Q-D---WVS---H-E------V--DEH.-------    504
CPZGAB      --K.--D-----P-------------...VS.T---------D-E---------NC------F---H.-------    506
CPZANT      V-S.--DR-QM-R------E--KQ-XXX-Q-KIE.-Y--Q-GLP-K-T-----S---------NEG.-L--A--    511
A_ROD       .-MT---E-QW--L--A--E---I-...-S--QE.-H--QEE-E-E-TV--DQEN----K-H--E..-I--V--    536
B_EHOACG    .-MT---E-QW--L--A-FQ--KI-...-E--QE.-S--KEGVP-E-TV--NLAN----K-H-GD..-I--V--    555
D_MM251     .-MT---E-QW--M--A-YE--KI-...-S--QE.-C--QEG-P-E-TVI-SQDN--S-K-H--D..-I--V--    556
STM_STM     .-MA---E-QW--M--A-Y---KI-...-S--QE.-Q--RED-P-E-TVV-DQDN--S-K-H--D..-I--V--    541
VER_AGM3    K-N.-L-T-TW-----A-Y---K--...--T-QE.-T--K-GRPIR-AV--LEG---S--FK--G..QV--V--    548
GRI_AGM677  K-N.-L-EIVW-----A-YKN-QG-...VQETQE.-T----L-E---TV----E------FT--G..AV--V-R    547
SAB_SAB1C   ARP.---I-QW--------E---Q-...-R-KQQ.-Q----ALP-R-KVL-L-D---G-----PEN.-I--V--    540
SYK_SYK     --Q.----LT---A--A--E---Q-...--E-QA.-S----K-P-E-H-T-L-SQ--G-M-K--QKGPP-I---    542
```

```
most-likely YARMKGAHTNDVKQLAEAVQKIGKESIVIWGKTPKFKLPIQKETWEQWWTEYWQATWIPEWEFVNTPPLVKLWY    570
A_U455      ---KRS---------T-V---VST--------I---R---------A--M------------------------    581
B_HXB2R     ----R----------T------TT---------------------T--------------------------    582
D_ELI       ----R---------------R-ST-------R----R---------T--A------------------------    581
O_ANT70     -T-Q-AS----IR----VI--VSQ---I----L------VTR----T--AD-------------S----I----    578
O_MVP5180   ---Q-AS----IR----V---VSQ-A------L---R--VTR----T--A--------------S----I----    578
CPZGAB      ---QRS-----IR---------AT------------R--V---S--A--A-------------I----------    580
CPZANT      ---PT-T---E-R---GV-----L-------EV---Q---TR---DA--SD------------------IR---    585
A_ROD       --KV-NT---GIRL--QV-------AL----RI---H--VER-I-----DN---V----D-D--S-----R-AF    610
B_EHOACG    --KV-NT---G-RL--HV-------AL----EI-M-H--VER---D----D---V------D--S----IR-A-    629
D_MM251     F-KI-NT---G-RL--HVI------A-----QV---H--VERDV------D---V------D-IS-----R-VF    630
STM_STM     F-KI-NT---G-RL--HVI------A-----QI---H--VER-V-----AD---V------D--S-----R-VF    615
VER_AGM3    -TKQ-NT---EFRV--GL---LC---L----EL-VLE---ER-V-----AD---VS---D----S---------    622
GRI_AGM677  --KQRET----LRT--HL----C--ALT---RL-RVQ--VD-K--DM--QD---VS--------S--L------    621
SAB_SAB1C   --KI-T----ELRM--GL-------------QI-IME--VER-L-----SD---V-------M-S--Q-IR---    614
SYK_SYK     T-KTFA--S--YQS--QLLN---IQ-LWY---V-T-H--VKR-E--K---D------V--VK-IS-----RWY-    616

                    p51 \/
most-likely QLEKEPIVGAETFYVDGAANRETKLGKAGYVTDRGRQKVVSLEDTTNQKTELQAIHLALQDSGLEVNIVTDSQY    644
A_U455      ----D--A----------------------------------TE--------H----------S----------    655
B_HXB2R     -------------------------------N-------T-T------------Y------------------    656
D_ELI       -------I--------------------------------P-T------------N------------------    655
O_ANT70     R--S---M----Y-------------------EQ-K--IIK-DE-----A--M--L------KET---------    652
O_MVP5180   ---T-----------------N----------EQ-K-NIIK--E-----A--M-VLI-----KEQ---------    652
CPZGAB      S--T---PTTD-Y-----------T--------K-K--II---N----QA--K-LL------DQQ---------    654
CPZANT      N-LAD--PE------------NSQ------------SR-KH-QK----QA-----LM--E--TGP---------    659
A_ROD       N-VGD--P------T--SC--QS-E----------KD--KK--Q----QA--E-FAM--T---PK---IV----    684
B_EHOACG    N-V-D-LE-V--Y-T--SC-KAS-E----------KD--KP--Q----QA--E-FA-------PQ---IV----    703
D_MM251     N-V-D--E-E--Y-T--SC-KQS-E-----I----KD--KV--Q----QA--E-FLM--T---PKT--IV----    704
STM_STM     N-V---LE-T-------SC--QS-E-----I----KN--KA--Q----QA--E-FAM--A---PKA---V----    689
VER_AGM3    T-T----PKEDVY-----C--NSRE-----I-QY-K-R-EK--N----QA--M--KM--E---PN---------    696
GRI_AGM677  S-V----K-EDVY------SKV--------LSE--KSRIRE--N----QA--T-VKM--E---EN---------    695
SAB_SAB1C   K-V-D--P-EAVY--------NS-E-----L----D----A--N-----A--E--L---R---SK---I-----    688
SYK_SYK     N-VP---PE-V----------DS-T-N----ASD-T-R-QY--Q----QA--EGLLM-----KDK----V----    690

most-likely ALGIIQAQPDESESELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSQGIRKVLFLEGIEPAQEEHEKY    718
A_U455      ----------R----I------K--E------S-------------------S--------D--DK---D----    729
B_HXB2R     ----------Q----------------------------------------A--------D--DK--D-----    730
D_ELI       ----------K-------------------------------------------------D--DK--------    729
O_ANT70     ---V-SS--TQ---PI-Q----E-T---Q---T------------KI-----KD--R-------DQ---D----    726
O_MVP5180   V----SS--TQ-D-PI-Q----E-T---R---T------------KI-----KD--R-------DQ---D----    726
CPZGAB      V-----S---H-----------E------I--S-------------------A--------D--DR------R-    728
CPZANT      ---VL-GT--Q---P--EE--QK---R-QI--S------------------------Q-------DK---D-D--    733
A_ROD       VM--SAS--T----KI------EM----AI-V------------QE--H-------Q-----K-----------    758
B_EHOACG    VM--VA---T-T--PI-RE---EM-----I-VG-------L---QE--H-------QI----K-----------    777
D_MM251     VM---TGC-T----R-------EM---SEI-V------------QEI-H-------Q-----K--------D--    778
STM_STM     VM---TG--T----K-------EM----AI-V------------QE--H-------Q-----K-----------    763
VER_AGM3    -M--LT---TQ-D-P-IE---ALMVQ-HQI--Q----D-------EI-------M--I----K--E------R-    770
GRI_AGM677  VMN-LT-C-Q--N-P--E---QA-M--RQ---Q-----------TEI-----K---QI---DR--E--DD-A--    769
SAB_SAB1C   -M---AGE-T--DNNI-Q----E-----A--I--------V----EI---------Q----DR--E-----D--    762
SYK_SYK     SY--LMTC-TNT-HPI-E---QEA----AI-VT------------A------K----I----R-PQ---D--R-    764
```

HIV1 POL

```
most-likely HSNWKELASKFGLPPLVAKEIVASCDKCQQKGEAIHGQVDASPGIWQMDCTHLEGKIIIVAVHVASGFIEAEVI    792
A_U455      -C--RAM--D-N---V---------N---L----M-----C------L--------V-L--------Y------    803
B_HXB2R     ----RAM--D-N---V-------------L----M-----C------L--------V-L--------Y------    804
D_ELI       -N--RAM--D-N---V-------------L----M-----C------L--------V-L--------Y------    803
O_ANT70     -----A---E-----V-----I---P--HI----------C--EV--I----M--------------------    800
O_MVP5180   ----RA---D-----I-----I---P--HI----T-----Y--E----------------------D-------    800
CPZGAB      -----AM--D-N---I-------H-----V----M-----C------V--------V----------Y------    802
CPZANT      ----RS--DEYN---I-----I-Q----HV----R-----C------V--------V-------S--------M    807
A_ROD       ---V---SH---I-N---RQ--N--AQ------------N-EL-T-----------------------------    832
B_EHOACG    -N-V---VH---I-Q---RQ--N----------------NSEL-T------------V-----------------    851
D_MM251     ---V---VF-----RI--RQ--DT----H----------NSDL-T------------V----------------    852
STM_STM     ---V---VF---I-R----Q--DT----H----------N-EL-T-----------------------------    837
VER_AGM3    -N--RN--DTY---QI-------M-P---I---PV---------V-----------V-----------------    844
GRI_AGM677  -N--RSMVQE----NI-------A-P---IR--PK-------IET-----------V-----------------    843
SAB_SAB1C   -A--RSMQQE----AI-------A-P---I---SV---------V-----------------------------    836
SYK_SYK     ---MEY-RQE-H--RQ---A-IQQ-P---NR--PK-----VDIYN-------E---V-C---NT---Y--TKIL    838

most-likely PQETGRETALFLLKLASRWPITHLHTDNGANFTSQEVKAACWWAGIEQTFGVPYNPQSQGVVESMNKHLKKIIG    866
A_U455      -A---Q---Y-I----G---VKVI-----S----AA---V----N-Q-E--I---------------E------    877
B_HXB2R     -A---Q---Y------G---VKTI-----S---GAT-R--------K-E--I---------------E------    878
D_ELI       -A---Q---Y------G---VKVV-----S----AA----------K-E--I---------------E------    877
O_ANT70     -A---Q---Y------A---VKVI-----P----TTM-------N-QHE--I-----------A---E--S--Q    874
O_MVP5180   -A---Q---Y------A---VKVI-----P----AAM------T--QHE--I-----------A---E--S--Q    874
CPZGAB      -A---Q---Y------G---VKTI-----P----AA--------D-K-E--I------------L--E------    876
CPZANT      AD---KS--Y----------VKTI----------AA--------N-Q-E--I---------------Q--Q---    881
A_ROD       ---S--Q-------------------------------MVA--I----S--------------A--H---NQ-S    906
B_EHOACG    ------Q----------------------------D--M-A--I------------E------A--H---NQ-D    925
D_MM251     ------Q---------G---------------A-----MVA------H---------------A--H---NQ-D    926
STM_STM     ------Q-------------V-----------------MVA----------------------A--H---TQ-D    911
VER_AGM3    -R---K---K----IL------Q------P-------A-M---GK--H-T----------SI-----Q--E---    918
GRI_AGM677  -R---K---H-----LA---VK-------P-----N-A-V---GN--H-T-I--------S-----RQ--E--S    917
SAB_SAB1C   -----KA--H------------Q------T-----Q-A-I---GK--H-------------------Q--E---    910
SYK_SYK     KR---D------MQI------KQI-----P--V-DKF------C---H-T-I--------I---K-RY--EA-S    912

most-likely QIRDQAEHLETAVQMAVHIHNFKRKGGIGGYTAGERIINIIATDIQTIQL.QKQISKIQNFRVYYREGRDPLWK    939
A_U455      -V-E-----K-------F-------------S------D--------KE-.---------------DS---I--    950
B_HXB2R     -V-------K-------F-------------S-----VD--------KE-.----T----------DS-N----    951
D_ELI       -V-------K-------F------RR-----S------D--------KE-.----I----------DS---I--    950
O_ANT70     -V-------R-------FV-------------------D-L-SQ---TE-.----L-                    930
O_MVP5180   -V-------K-------FV-----------------L-D-L-SQ---TE-.----L--N-------DS---I--    947
CPZGAB      -V-------K-------F--------------------D--------SE-.----L-V-K------DS---I--    949
CPZANT      -------Q-K---V------------------P-Q--LD-L------T--.-N--L---Q---H--DS---V--    954
A_ROD       R--E--NTI--I-L--I-CM----R----DM-PS--L--M-T-E.-E--FL-AKN--LKD----F-----Q---    979
B_EHOACG    R-----VSI--V-L--T-CM----R----DM-PA---V-M-T-E.-E--FL-TKNL-F------------Q---    998
D_MM251     R--E--NSV--I-L----CM----R----DM-PA--L--M-T-E.-E--FQ-SKN--FK-----------Q---    999
STM_STM     R-----NTV--V-L----CM---KR--L-DM-PA--LV-M-T-E.-E--FQ-SKN--FK-----------Q---    984
VER_AGM3    K---DCQYT----L--C-------------L-PA--L--M-T-Q.LEL-HL-TK-Q--L------------V--    991
GRI_AGM677  ----DC-R--------T-------------ISSA--LV-MLT-Q.LELNTL-N--Q--L--K---------V--    990
SAB_SAB1C   ----D--R-----I-----------------S-A--L----H-ELE-KT-.-QK-----------------V--    983
SYK_SYK     ----DVT--Q---A--TF-L----------ISP---Y--MLY-E.LQL-Q.NTTSP-FS-------Q-KNE.--    983
```

```
                                      vif cds ->
most-likely GPAKLLWKGEGAVVIQ.DGSDIKVVPRRKAKIIRDYGKQMAGDDCV............D                  985
A_U455      ----------------.-N--------------------------M...AGRQD..E.-                 1002
B_HXB2R     ----------------.-N----------------------------...ASRQD..E.-                 1003
D_ELI       ----------------.-K----------V----------------...ASRQD..E.-                 1002
O_MVP5180   ---Q------------.-KG----------------------T-SM...ANRQT..E.SESMEQPGEIP        1009
CPZGAB      ---T------------.-QGEL-------------------------...ASRQN..E.-KHGTAW           1007
CPZANT      ---Q-----------                                                               969
A_ROD       --GE---------LVK.V-T---II------------GRQEM-SGSHLEGAREDG.EMA                 1036
B_EHOACG    --GD---------I-K.V-TE---I---------N--GGKEL-CSADVEDTMQAR.EVAQSN               1058
D_MM251     --GE---------ILK.V-T-------------K---GGKEV-SSSHMEDTGEAR.EVA                 1056
STM_STM     --GE---------IVK.V-T-------------K---GGKEV-SGSHLEDTREAG.EVA                 1041
VER_AGM3    --GQ-I---------K.G-VEL-EY--------K--EPRKRMG-ESNLEGAGGA..D.N                 1046
GRI_AGM677  ---R-I---------K.E-E-------------K---ERKTM-SEGSMEGVREANKQMEGDSDLQDQE         1057
SAB_SAB1C   -----I----------.EQGEL-TI--------K----ALDSQAPL.EGNGRTAG.EV-                 1039
SYK_SYK     ---R----------V-TEEG--FA---------T-H-ER-DSGSH-.....ENDP.KT-                 1036
```

```
      **COBBLER sequence from MOTIF**
>POL   CPZGAB, with embedded consensus blocks
ffrerlafpqrearqlcaeqnrtngptdrelwvpggreepgeergreqsistnlpqitlw
qrplipvkvegqlcealldtgaddtvieriqlqglwkpkmiggiggfikvkqfdnvhiei
egrkvvgtvlvgptpvniigrniltqlgctlvfpissietvpvklkpgmdgpkvkqwpls
aekikalteicqemekegkiskigpeNPYNTPVFAIKKKDKTKWRMLVDFRELNKRTQDF
WEVQLGIPHPAGLKKKKSVTVldvgdayfscpldkDFRKYTAFTIPSVNNETPGVRYQYN
VLPQGWKGSPAIFQSTMTKILEPFReknpditiyqymddlyvgsdleidqHRRMVEQLRE
HLLKWGFETPDKKHQKEPPFQWMGYELHPDKWTIQKIqlpekevWTVNDIQKLVGKLNWA
SQIYPGIKTKHLCKLIRGTKSLTEVVPWTEEAEAELEEnreivstpvhgvyydpdkelia
eiqkqgncqwtyqifqephknlktgkyarqrsahtndirqLAEVVQKIGKESIVIWGKIP
KFHLPVERETWEQWWTDYWQATWIPEWEFVSTpplvklwysletepipttdtyyvdgaan
retktgkAGYVTDRGKQKVKKLEQTTNQQAELEAILMALQDSGPKVNIVTDSQYAMGIIq
sqpdhseselvnqiieelIKKEKIYLAWVPAHKGIGGNEEVDKLVSQGIRQVLFLEGIDE
AQEEHEKYHSNWRamasdFGLPPIVAKEIVASCPKCQIKGEAMHGQVDCSPGVWQMDCTH
LEGKVIIVAvhvaSGFIEAEVIPQETGQETAYFLLKLASRWPVKHIHTDNGPNFTSQEVK
AACWWadikqefGIPYNPQSQGVVESMNKHLKEIIGQIRDQAEHLETAVLMAVHIHNFKR
KGGIGGytageriidiiatdiqtselqkqilkvqkfrvyyrdsrdpiwkgpatllwkgeg
avviqdqgelkvvprrkakiirdygkqmagddcvasrqnedkhgtaw
```

HIV1 POL CONSENSUS

```
CONSENSUS-A   FFRE.NLAFQQGEAR?F............SSE..QT??NS?TSR?LWDGG?D??.L?...???G?E?..Q     35
CONSENSUS-B   ----.d---p--k--e-???????????---..--Ra--p-r-E-qVw-r-nnS-S???-EA-adr..-     49
CONSENSUS-D   ----.d---P--K-GEl.............---..--RA--P---E-RVW-r-.NP-S....eT-A-R..-     48
CONSENSUS-O   ---?.?--SGGH---QL............CA-..TS-PI-P-?.....--GSE.....GT-ES?---G??     35
CONSENSUS-U   ----.----P--K--E-............P--..--RA--P---E-RVW-G-K.T-S....ET-A-R..-     48
CONSENSUS-CPZ ----??????????--L............CA-????---?--?---????-?--??-....--?-?-???     13

                    protease
                      \/           <- gag cds end
CONSENSUS-A   G?..??SF?FPQITLWQRPLVTV?I?GQLIEALLDTGADDTVLEDINLPGKWKPK?IGGIGGFIKVRQYD     96
CONSENSUS-B   -...tV--s-------------ik-g---K--------------eM----r----M--------------    116
CONSENSUS-D   -...TV--n-------------IK-G---K--------------Em---------M--------------    115
CONSENSUS-O   R...A-??CL---P--D--I--A-VG-H-C-?-----------NN-Q-E-?-?--M----------KE-?     94
CONSENSUS-U   -...IV--S----------V---RVG---K--------------E----------M--------------    115
CONSENSUS-CPZ ?-??---?-?-?-?-----???--?---?C?-----------?-?-?-Q-?-------------?-?-??     55

                                          protease \/ p66, p51
CONSENSUS-A   QILIEICGKK?IGTVLVGPTPVNIIGRNMLTQIGCTLNFPISPIETVPVKLKP?MDGPKVKQWPLTEEKI    164
CONSENSUS-B   --------H-A-----------------L------------------------G----------------    186
CONSENSUS-D   --------?-A-----------------L------------------------G----------------    184
CONSENSUS-O   NVTV-??-?EVQ------------?---I--GL-----------AP-------G-----------S?---    159
CONSENSUS-U   ----------A---I-------------------------------------G----R-----------    185
CONSENSUS-CPZ ?V?-?-??R?V??---?--?--------?-??L----?----??--------?G----?------S?---    106

                     | M41L                D67N |  | K70R
CONSENSUS-A   KALT?IC?EMEKEGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPH?AGLK    231
CONSENSUS-B   ---vE--T-------------------------------------------------------P----    256
CONSENSUS-D   ----E--T---------R----------I-----------------------------------P----    254
CONSENSUS-O   E---A--Q---Q-----R----------I-------?-----------------?----------PG---    227
CONSENSUS-U   ----E--KD-------L---------------------------N--------------------P----    255
CONSENSUS-CPZ ?--?E---??-?-?---?---?------I-------??--------------------?------P----    164

CONSENSUS-A   KKKSVTVLDVGDAYFSVPLD??FRKYTAFTIPS?NNETPG?RYQYNVLPQGWKGSP?IFQ?SMTKILEPF    295
CONSENSUS-B   --------------------kd-----------i------i---------------A---s---------    326
CONSENSUS-D   --------------------eD-----------I------I---------------A---S---------    324
CONSENSUS-O   Q?Q-------------C---PD-----------V----------------------A---S------D--    295
CONSENSUS-U   --------------------ED-----------I------I---------------A---S---------    325
CONSENSUS-CPZ ?---------------?----D---------------------?-----------------------?--    225

                  polymerase motif
                        /<-->/                    T215Y |   | K219Q
CONSENSUS-A   R???P?IVIYQYMDDLYVGSDLEIGQHRAKIEELR?HLL?WGF?TPDKKHQKEPPFLWMGYELHPDKWTV    358
CONSENSUS-B   -KQN-d----------------------t------q---r---t--------------------------    396
CONSENSUS-D   -KQN-E----------------------T------E---R---T--------------------------    394
CONSENSUS-O   ---N-E?E-?--?---------PL?E--KR?-L--E--YQ---T--------------------------    358
CONSENSUS-U   -TKN-E----------------------T------E---K---T--------------------------    395
CONSENSUS-CPZ --K---??-------------?--??--???-?--Q--?---?-------------?-------------    278

CONSENSUS-A   QP??LPEKDSWTVNDIQKLVGKLNWASQIYAGIK?KQLC?LLRGAKALTDIV?LTEEAELELAENREI..    421
CONSENSUS-B   --Iv------------------------------V----k----t----Evip---------------..    464
CONSENSUS-D   -sIk----E---------------------p---Vr---K----T----EViP---------------..    462
CONSENSUS-O   -?IQ--?-?V---------?----------Q--RV?E--K-I--T-S--EV-P-S?------E----?..    419
CONSENSUS-U   --IQ--D-E---------------------P---V----K------------P--A------------..    463
CONSENSUS-CPZ -?I----???---------?----------P----------I--?-?---?--?-?------?--??-??    329

CONSENSUS-A   .LK?PVHGVYYDP?KDLVAE?QKQGQDQWTYQIYQEPFKNLKTGKYA?KRSAHTNDVKQLTEVVQKV??E    484
CONSENSUS-B   .--e---------s---i--i-----g--------------------rm-G-----------A---iat-    533
CONSENSUS-D   .--E---------S---I--i----hG--------------------Rm-G---------a-a---IsT-    531
CONSENSUS-O   .----------Q-D---WV?I--?-??-----?--?EH--------?RQKAS----IR--A--?---SQ-    479
CONSENSUS-U   .--E---------S---I--I-----G---------QY---------RIK------------A---IAQ-    532
CONSENSUS-CPZ ???-???-?--?--??-?-?I----??------?-???-?--?----R????---??R--A??---I---    367
```

```
                                                                p51 \/
CONSENSUS-A   SIVIWGK?PKFRLPIQ?ETWE?WWMEYWQATWIPEWEFVNTPPLVKLWYQLEKDPI?GAETFYVDGAANR    550
CONSENSUS-B   -------t---k----K-------t----------------------------e--v-------------    602
CONSENSUS-D   -------T--------K----T--?----------------------------E--I-------------    600
CONSENSUS-O   ?-?----L---?--VTR----T--A?-------------S----I----?--?E-------?--------    541
CONSENSUS-U   -------T--------K----A--T---------------------------TE--V-------------    602
CONSENSUS-CPZ ------?----?--??--?-?A--??------------?-----??---?-???--P???-?--------    416

CONSENSUS-A   ETK?GKAGYVTDRGRQKVVSLTETTNQKTELHAIHLALQDSGSEVNIVTDSQYALGIIQAQPDRSESE?V    618
CONSENSUS-B   ---l------------------d--------q----------l--------------------k----l-    672
CONSENSUS-D   ---L---------------Pf-D--------Q--N-------L--------------------K----L-    670
CONSENSUS-O   ?--L-------EQ-K-?IIK-?------A--M-?L?-----KE?---------?--?-SS--TQ-?-PI-    602
CONSENSUS-U   ---K---------------------------Q------------------------------K----I-    672
CONSENSUS-CPZ ???---------?-??????-??----QA--?-?L?--?--???---------?--??-??--?---?L-    459

CONSENSUS-A   NQIIEKLI?K?KVYLSWVPAHKGIGGNEQVDKLVS?GIRKVLFLDGIDKAQE?HE?YH?NW?AMASDFNL    681
CONSENSUS-B   s----q--K-E----a-------------------a----------------e--K--s--r--------    742
CONSENSUS-D   s----Q--K-E----A-------------------Q----------------E--K--N--R--------    740
CONSENSUS-O   Q----E-TK-E?---T------------KI-----KD--R----E---Q---D--K--S----L--?-G-    669
CONSENSUS-U   -----Q--Q-D------------------------S----------------E--K--S--R--------    742
CONSENSUS-CPZ ??--??--K?E?I--------------------------?----?---?-----?---S---??-???--    510

CONSENSUS-A   PP?VAKEIVASC?KCQ?KGEAMHGQVDC?PGIWQLDCTHLEGK?ILVAVHVASGYIEAEVIPAETGQETA    746
CONSENSUS-B   --v---------d---L-----------S--------------v--------------------------    812
CONSENSUS-D   --V---------D---L-----------S--------------V--------------------------    810
CONSENSUS-O   --------I---P--HI----?-----?S-E?--?----?---I-I-------?F---------------    732
CONSENSUS-U   --I---------D---L-----------S--------------I-I------------------------    812
CONSENSUS-CPZ --I-----?-?-D--?V----?------S-----V--------V-I-----?--?-----???---??--    569

CONSENSUS-A   YFILKLAGRWPVKVIHTDNGSNFTSAAVKA?CWWAN??QEFGIPYNPQSQGVVESMNKELKKIIGQVR?Q    812
CONSENSUS-B   --l----------t-----------tT---A----GiK------------------------------D-    882
CONSENSUS-D   --l-----------V---------------A----GIK------------------------------D-    880
CONSENSUS-O   --L----A------------P----??M--A---??IQH---------------A------S--Q---D-    798
CONSENSUS-U   --------------V---------------A-----IK-----------------------------E-    882
CONSENSUS-CPZ --L----?-----T------?---------A----?I------------------?--?--?----?-D-    631

CONSENSUS-A   AEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIIDIIA?DIQTKELQKQI?KIQNFRVYYRDSRDPIWKGP    880
CONSENSUS-B   --------------------------------v----t-----------T---------------l----    952
CONSENSUS-D   -------------------------------------T-----------i--------------------    950
CONSENSUS-O   ----?--------V------------T----?---L-SQ---T------L-?N-----------------    865
CONSENSUS-U   ----------------------------------M--T-----------T-----------N--------    952
CONSENSUS-CPZ --?-----?---?-------------T?-?--?--?-T----??--?--L-?-?---?-------?----    687

                                     vif cds ->
CONSENSUS-A   AKLLWKGEGAVVIQDNSDIKVVPRRKAKIIRDYGKQMAGDDC?AGRQDED                       929
CONSENSUS-B   ------------------------------------------V-s-----                      1002
CONSENSUS-D   --------------------------V---------------V-S-----                      1000
CONSENSUS-O   -Q-------------KG----------------------T-SM-N--T-SESMEQPGEIP              925
CONSENSUS-U   ------------------------------------------V--G----KHGTAW                 1008
CONSENSUS-CPZ -?-------------QGEL-----------------------V-S--N--KHGTAW                 742
```